Захаров А. И., аспирант
Научный руководитель Лидский Э.А, проф., д-р техн. наук

АВТОМАТИЗАЦИЯ ПРОЦЕССА ФОРМИРОВАНИЯ ПЛАТЕЖНОЙ МАТРИЦЫ

Необходимость автоматизации процесса формирования платежной матрицы в игре защита-нападение обусловлена двумя основными причинами. Во-первых, при работе с большими выборками из базы данных ICAT объем информации, которую необходимо обрабатывать, слишком велик. Во-вторых, используемый в работе подход с назначением предпочтительности требует многократного пересчета и минимизации самой платежной матрицы, и для получения конечного результата требуется выполнить большой объем вычислений.

Процесс автоматизации подразумевает создание алгоритмов для выполнения каждого этапа работы. Целью настоящей статьи является обзор используемых методов.

1.Создание словаря ключевых слов

Процесс декомпозиции базы данных ICAT может продолжаться до тех пор, пока остается возможность принимать решение о предпочтительности того или иного варианта. В дальнейшем вероятность выбора варианта атаки можно оценить, предложив вместо сокращения подмножества на очередном шаге декомпозиции выявлять только относительные частоты повторения намечаемых ключей. Под намечаемыми ключами здесь понимаются не основные атрибуты базы данных ICAT, а дополнительные данные, которые можно извлечь из поля Description. Содержание поля Description можно считать формализованным, поскольку выдерживается, во-первых, морфологическая структура излагаемой информации, во-вторых, используются одни и те же термины для описания одних и тех же объектов и событий.

Задача поиска ключевых слов основана на механизме индексирования, используемом в глобальной сети Интернет. Этот подход можно считать полностью унифицированным, поскольку в современном Интернете единственным способом поиска информации является использование поисковых машин, которые, в свою очередь, занимаются индексированием всего множества информационных источников сети. Необходимо отметить, что метод индексирования очень хорошо подходит для задачи поиска ключевых слов в описательной части базы данных ICAT. Нужно лишь определиться, какой разновидностью алгоритма индексирования воспользоваться. Данный вопрос может являться предметом отдельной научной работы, поскольку необходимо не только изучить существующие способы индексирования, но и предложить методику выбора оптимального из них, опираясь на определенный набор критериев. Такая задача выходит за рамки настоящей работы, однако основные моменты все же отметим.

Большинство специализированных систем индексирования основаны на каких-либо разновидностях семантического анализа. Существуют целые программные комплексы, решающие задачи распознавания текста и разложения его на морфологические составляющие. Для индексирования web документов в Интернете, как правило, столь серьезные алгоритмы не нужны. Для поиска ключевых слов, по которым обычно осуществляется поиск информации, достаточно выбрать наиболее часто встречающиеся слова и словосочетания, поскольку на основании частоты, с которой встречаются в тексте те или иные группы слов и выражений, можно судить о тематике ресурса в целом. Для нашей задачи такой подход является наиболее подходящим, поскольку за счет выявления слов и фраз с максимальной частотой повторения в пределах всего объема области Description становится возможным выделить наиболее часто повторяющиеся словосочетания и избрать их на роль ключей. Алгоритмы, решающие такую задачу, очень просты и работают по принципу удаления связующих частей речи.

В настоящей работе используется алгоритм, автор которого – Andy Hoskinson – является одним из авторов серии книг по Интернет программированию. Разработка является интеллектуальной собственностью автора, однако она свободно размещена на сайте автора www.hoskinson.net и доступна к использованию. Алгоритм реализован в виде исполняемого скрипта, который, получая на входе некоторый объем информации, а также список стоп-слов, выдает на выходе перечень наиболее часто встречающихся словосочетаний и частоту их повторения. В нашей задаче наиболее часто встречающиеся слова претендуют на роль ключей, поскольку являются унифицированным описанием объектов или действий, а следовательно, могут быть интерпретированы вполне однозначно.

2.Использование ключей

Для разрешения игровой задачи используется платежная матрица, элементы которой представляют собой значение выигрыша/проигрыша сторон при выборе сторонами определенного варианта атаки/защиты. Элемент платежной матрицы представляется в следующем виде:

$$a_{ij} = \sum_{k=1}^{K_i} c(i,j)(b_k(j) - \bar{b}_k(i))^2,$$

где в роли $\bar{b}_k$ и b_k выступают ключи, сгенерированные на предыдущем этапе работы. По окончании декомпозиции с использованием основных атрибутов базы данных ни нападение, ни защита не видят предпочтительных вариантов среди имеющихся. Использование дополнительных ключевых слов позволяет оценить степень взаимного перекрытия каждой пары "атака-защита", на основании сравнения информационных контекстов, присутствующих в характеристике каждой уязвимости.

Для того чтобы выразить в численном виде степень перекрытия отдельно взятой пары вариантов, необходимо подсчитать количество не совпавших ключей. Поскольку атакующий игрок всегда является ведущим, подсчет ведется относительно него.

Число получившихся ключей у разных описаний может существенно варьироваться, но в нашем примере оно получилось равным.

Для того чтобы иметь возможность сравнивать ключи, понадобится работа с массивами. Учитывая, что каждое описание уязвимости в базе данных ICAT имеет свой глобальный идентификатор, для сравнения каждой пары описаний будут создаваться два массива M1[i][k] и M2[j][k]. Каждый элемент массива, фактически, представляет собой элемент $b_k(i)$, где i – номер описания, а k – номер ключа в описании.

Производится непосредственное сравнение получившихся массивов с целью определения степени перекрытия намеченных ключей. Сравнение производится относительно ведущего игрока, которым выступает атакующий. Необходимо отметить, что обычное попарное сравнение ключей здесь неприменимо, поскольку позиция одних и тех же ключей может варьироваться в зависимости от конкретного описания.

Использование ключевых слов из описательного поля Description позволяет получить информационный контекст действия уязвимости, а изучение перекрытия ключевых слов позволит оценить насколько избранный вариант защиты эффективен против определенного варианта атаки. Это вполне конкретное сравнение, поскольку сравнивается информация и по механизмам проявления уязвимостей, и по подверженным программным компонентам.